Documento de Trabajo 2003-03

Facultad de Ciencias Económicas y Empresariales

Universidad de Zaragoza

Selection of the Informative Base in ARMA-GARCH Models

Laura Muñoz, Pilar Olave and Manuel Salvador

Department of Statistics Methods, University of Zaragoza

Abstract. In this paper we consider the selection of the information set in ARMA-GARCH models using the methodology proposed in Muñoz et al. (2001) based on ideas of Phillips (1996). To that end, we analyse the performance of some selection criteria asymptotically equivalent to the Bayes factor and propose a method to build approximated Bayesian forecast intervals.

Keywords: ARMA, GARCH, Prediction , Model Selection, Information Set Selection.

JEL Classification: C11, C22, C51

Address: Laura Muñoz Garatachea. Departamento de Métodos Estadísticos. Facultad de Ciencias Económicas y Empresariales. Universidad de Zaragoza. Gran Vía, 2. 50005 Zaragoza. Spain. E-mail: lmunoz@unizar.es

1. Introduction

In the analysis of financial series, the agents' interest lies in obtaining reliable forecasts in the short and medium term, above all in the presence of unexpected news (innovations). This is the case because such circumstances lead to forecasts which have a limited informative content due to the existence of inputs that substantially modify the model and the information set.

From among the first studies devoted to this type of problem, attention should be drawn to those of Baillie and Bollerslev (1992) and Bera and Higgins (1993), who indicated the great importance of evaluating the conditional variance of the forecast errors. That is to say, particular care should be taken when calculating the uncertainty associated to the conditional variances in the various stages, i.e. the short, medium and long term. Miguel and Olave (2002) have demonstrated that this, in turn, points us towards another aspect where care must similarly be taken, namely the correct selection of the most appropiate information set, in such a way that the uncertainty associated to the construction of confidence intervals does not lead to large indeterminations in the medium term forecast.

Phillips and Ploberger (1994) and Phillips (1995, 1996) have proposed a procedure that considers the model selection problem in a very general Bayesian context in which it is assumed that the Data Generating Process (DGP) does not have to belong to the family of models proposed by the analyst and, furthermore, can evolve over time. Phillips (1996) presents an asymptotic approximation to the prior forecast density (see Theorem 1 below), which he later uses to approximate the Bayes factor, proposing a model selection criterion, the PIC (Posterior Information Criterion) criterion that is based on such an approximation. This criterion naturally combines the goodness of fit and the parsimony of the proposed models when selecting the optimal model.

Using this approach, it is also possible to select the most appropiate information sets in order to estimate the parameters of the proposed models and to forecast the future behaviour of the series being analysed. For that purpose, Phillips (1996) uses an asymptotic approximation to the posterior forecast distribution, on the basis of which he presents the PICF (Posterior Information Criterion for Forecasts) criterion which, in addition to the above, also takes into account the most appropiate information set to maximise the forecast density of the models.

Recently, Muñoz et al. (2001) have proposed a methodology to select the model and the information set in the heteroscedastic context proposed by Lejeune (1997) and give sufficient conditions for the weak and strong consistency of the criteria employed. Particular cases of this methodology are the PIC and the PICF criteria of Phillips (1996) and Phillips and Ploberger (1994) and the BIC criterion of Schwarz (1978).

In this paper the methodology of Muñoz et al. (2001) is applied to the statistical analysis of ARMA-GARCH models. These models have been extensively used in the analysis of financial series (see Bollerslev et al. (1992) and Bera and Higgins (1993)). In this context, the main objective of this paper is to propose a method to solve the problem of selecting the information set in such a way that the overvaluation of the persistence in volatility, higlighted in Lamoureux and Lastrapes (1990), does not affect the short and medium term forecasts. It is precisely these forecasts, which are of interest in Financial Economics. To that end, we study the behaviour in finite samples of various model selection criteria that are asymptotically equivalent to the partial Bayes factor under certain regularity conditions. Our aim here is to obtain the optimum information base of a model that evolves over time. From among the criteria analysed, we find that the BIC criterion and the PRED and RPRED criteria, which approximate the posterior forecasting density by way of simulation, show a better behaviour than the PICF criterion. This latter criterion tends to choose the smallest information sets when the DGP is less parsimonious and has not changed during the period being analysed.

Furthermore, we demonstrate how to make short and medium term apporximated forecasts taking into account the uncertainty associated with the estimation of the parameters of the model. In doing so, we make use of the Monte Carlo method and the asymptotic normality of the posterior distribution of the parameters of the model which are deduced from the results of Muñoz et al. (2001).

The rest of the paper is organised as follows. In Section 2 we establish the problem, describe the methodology employed and demostrate how to obtain, in an approximated form, the Bayesian forecast intervals by Monte Carlo. Section 3 is devoted to studying the behaviour of the criteria proposed in the earlier section in ARMA-GARCH models. Finally, Section 4 closes the paper with a review of the main conclusions. The mathematical demonstrations of the results are relegated to an Appendix.

2. The Problem

Let $(\mathbf{R}^\infty,\mathbf{B}(\mathbf{R}^\infty),P_o)$ be a complete probability space where $\mathbf{B}(\mathbf{R}^\infty)$ denotes the Borel σ-algebra, and let Y be a univariate stochastic process. In what follows P_o will be called the Data Generating Process of Y (DGP). Let $\{y_1,...,y_T\}$ be the observed values of Y in period $\{1,...,T\}$. Further, let **S** be the family of univariate ARMA-GARCH models given by the equation:

$$y_t = \mu + \sum_{i=1}^{nar}\gamma_i y_{t-i} + \sum_{i=1}^{nma}\theta_i\varepsilon_{t-i} + \varepsilon_t \quad t = 1,...,T \tag{2.1}$$

with $\varepsilon_t \mid \Phi_{t-1} \sim D(0,h_t)$ where Φ_t is the σ-algebra generated by $\{y_1,...,y_t\}$, $D(0,h_t)$ a probability distribution with 0 mean and variance h_t given by:

$$h_t = \omega + \sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\beta_j h_{t-j} \quad \text{where } h_t = 0 \text{ and } \varepsilon_t = 0 \quad t \le 0 \tag{2.2}$$

with nar, nma, p, q $\in\mathbf{N}\cup\{0\}$ and $\theta = (\mu, \gamma_1,..., \gamma_{nar}, \theta_1,...,\theta_{nma}, \omega, \alpha_1,...,\alpha_p, \beta_1,...,\beta_q)' \in\Theta\subseteq A$ a compact subset of A where A is given by:

A = $\{\theta$: $\mu\in\mathbf{R}$, $\gamma_k\in\mathbf{R}$, k=1,...,nar; $\theta_l\in\mathbf{R}$, l=1,...,nma; $\omega > 0$, $\alpha_i \ge 0$, i=1,...,p ; $\beta_j \ge 0$, j=1,...,q, $\sum_{i=1}^{p}|\alpha_i| < 1$, $\sum_{j=1}^{q}|\beta_j| < 1$, and the roots of the polynomials 1 - $\sum_{i=1}^{nar}\gamma_i z^i$, 1 + $\sum_{i=1}^{nma}\theta_i z^i$ and 1 - $\sum_{l=1}^{q}\beta_l z^l$ are outside of the unit circle$\}$. (2.3)

If the conditional distribution D is normal and conditioning on the first t observations, the likelihood function of the model **S** in period $\{s_1, ..., s_2\}$ $1\le s_1<s_2\le T$ will be:

$$L_{s_1,s_2}(\theta) = \prod_{t=s_1}^{s_2}\left[\frac{1}{\sqrt{2\pi}}\frac{1}{h_t^{1/2}}\exp\left[-\frac{1}{2}\frac{\varepsilon_t^2}{h_t}\right]\right] \tag{2.4}$$

and its logarithm will be:

$$\ell_{s_1,s_2}(\theta) = -\frac{(s_2 - s_1 + 1)}{2}\log(2\pi) - \sum_{t=s_1}^{s_2}u_t(\theta) \tag{2.5}$$

with $u_t(\theta) = -\frac{1}{2}\log(h_t)$ - $\frac{\varepsilon_t^2}{2h_t}$.

Selecting the information set

Let us assume that $\exists 1 \leq t < T$ such that the DGP may have changed in period $\{1,...,t\}$ but not in $\{t+1,...,T\}$ and, in this period, $P_o \in \mathbf{S}$. The period $\{1,...,t\}$ will be called the *estimation period* and the period $\{t+1,...,T\}$ the *validation period.* The data sets $Y_e = \{y_1,...,y_t\}$ and $Y_v = \{y_{t+1},....,y_T\}$ will be called *estimation and validation sample*, respectively.

Our objective is to determine the most adequate subperiod of $\{s,...,T\}$ where $s \in \{1,...,t\}$ are used to estimate the DGP in order to forecast the future evolution of Y. To that end, a family of relevant information sets, $\mathbf{F} = \{\mathbf{F}_{s,t} = \mathbf{B}(\{y_s,...,y_t\}: s \in \{s_o ,..., s_f\}$ with $1 \leq s_0 < ... < s_f < t\}$, is considered. Our objective will be to choose $\mathbf{F}_{s_{opt},t} \in \mathbf{F}$ such that:

$$\mathrm{CRIT}(Y_v \mid \mathbf{F}_{s_{opt},t} ;\mathbf{S}) = \max_{s \in \{s_0,...,s_f\}} \mathrm{CRIT}(Y_v \mid \mathbf{F}_{s,t} ;\mathbf{S}) \tag{2.6}$$

where $\mathrm{CRIT}(Y_v \mid \mathbf{F}_{s,t} ;\mathbf{S})$ is a criterion, to be specified (see Section 2.2 below), which evaluates the predictive capacity of the information set $\mathbf{F}_{s,t}$ in order to explain the validation sample, Y_v.

Selection Criteria

The CRIT criterion should be consistent, i.e., so that for large validation samples and/or large information sets, the DGP could be consistently estimated in the validation period using the period $\{s_{opt},...,T\}$. Muñoz et al. (2001) propose using expressions that are asymptotically equivalent to the posterior predictive density given by:

$$m(Y_v \mid F_{s,t},\pi,\mathbf{S}) = \int_\Theta L_{t+1,T}(\theta)\pi(\theta \mid F_{s,t})d\theta \tag{2.7}$$

where $\pi(\theta \mid F_{s,t}) = \dfrac{L_{s,t}(\theta)\pi(\theta)}{\int_\Theta L_{s,t}(\theta)\pi(\theta)d\theta}$ is the posterior distribution of θ corresponding to a prior distribution π.

In order to specify these criteria, let $\hat{\theta}_{s,t}$ $(1 \leq s < t)$ be a quasi-maximum likelihood estimator (QMLE) of θ in the period $\{s,..., t\}$ using the likelihood function (2.4). In this paper the following criteria will be considered:

- *Bayesian Information Criterion for Forecasting (BICF)*: This is based on the BIC criterion of Schwarz (1978) and is given by:

$$\text{BICF}(Y_v \mid \mathbf{F}_{s,t};\mathbf{S}) = \left(\ell_{s,T}(\hat{\theta}_{s,T}) - \ell_{s,t}(\hat{\theta}_{s,t})\right) - \frac{p(S)}{2}(\ln(T\text{-}s)\text{-}\ln(t\text{-}s)) \qquad (2.8)$$

where p(S) is the number of parameters of the model **S**.

- *Posterior Information Criterion for Forecasting (PICF)*, proposed by Phillips (1996) and Phillips and Ploberger (1994). This is given by:

$$\text{PICF}(Y_v \mid \mathbf{F}_{s,t};\mathbf{S}) = \left(\ell_{s,T}(\hat{\theta}_{s,T}) - \ell_{s,t}(\hat{\theta}_{s,t})\right) - \frac{1}{2}\left(\ln\left(\left|B_{s,T}(\hat{\theta}_{s,T})\right|\right) - \ln\left(\left|B_{s,t}(\hat{\theta}_{s,t})\right|\right)\right) \qquad (2.9)$$

where $B_{s,t}(\theta)$ is the conditional information matrix of the model **S** given by :

$$B_{s,t}(\theta) = \sum_{i=s}^{t} E\left[\frac{du_i}{d\theta}\frac{du_i}{d\theta'}\middle|\boldsymbol{\Phi}_{i-1}\right] = \sum_{i=s}^{t}\left(\frac{1}{2h_i^2}\frac{dh_i}{d\theta}\frac{dh_i}{d\theta'} + \frac{1}{h_i}\frac{d\varepsilon_i}{d\theta}\frac{d\varepsilon_i}{d\theta'}\right) \qquad (2.10)$$

where $\left\{\frac{dh_t}{d\theta}, \frac{d\varepsilon_t}{d\theta}; t = 1,\dots,T\right\}$ are given by the recursive expressions:

$$\frac{dh_t}{d\theta} = D_t + 2\sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}\frac{d\varepsilon_{t-i}}{d\theta} + \sum_{j=1}^{q}\beta_j\frac{dh_{t-j}}{d\theta} \qquad (2.11)$$

$$\frac{d\varepsilon_t}{d\theta} = A_t - \sum_{i=1}^{nma}\theta_i\frac{d\varepsilon_{t-i}}{d\theta} \qquad (2.12)$$

with $\frac{dh_t}{d\theta} = \frac{d\varepsilon_t}{d\theta} = \mathbf{0}$ if $t \le 0$ and A_t, D_t given by:

$$A_t = (-1, -y_{t-1}, \dots, -y_{t-nar}, -\varepsilon_{t-1}, \dots, -\varepsilon_{t-nma}, 0, \mathbf{0}_p, \mathbf{0}_q)' \qquad (2.13)$$

$$D_t = (0,0,\mathbf{0}_{nar},\mathbf{0}_{nma},1,\varepsilon_{t-1}^2, \dots, \varepsilon_{t-p}^2, h_{t-1},\dots,h_{t-q})' \qquad (2.14)$$

where $\mathbf{0}_p$ denotes the px1 null vector.

- *The predictive density criterion (PRED)*

This criterion is based on the asymptotic normality of the posterior distribution $\pi(\theta \mid \mathbf{F}_{s,t})$ for large information sets. It can be shown that, under some regularity conditions on the prior distribution π, $\pi(\theta \mid \mathbf{F}_{s,t}) \approx N_{p(S)}\left(\hat{\theta}_{s,t}, B_{s,t}^{-1}(\hat{\theta}_{s,t})\right)$ when t-s $\to\infty$ where $N_{p(S)}(\hat{\theta}_{s,t}, B_{s,t}^{-1}(\hat{\theta}_{s,t}))(\theta)$ denotes the density function of a normal distribution with mean vector $\hat{\theta}_{s,t}$ and covariance matrix $B_{s,t}^{-1}(\hat{\theta}_{s,t})$ [see Muñoz et al. (2001), corollary 3.2].

The predictive density criterion is given by:

$$\mathrm{PRED}(Y_v \mid \mathbf{F}_{s,t},\mathbf{S}) = \int_{R^{p(S)}} L_{t+1,T}(\theta) N_{p(S)}(\hat{\theta}_{s,t}, B_{s,t}^{-1}(\hat{\theta}_{s,t}))(\theta)d\theta \tag{2.15}$$

This expression is usually analitically intractable, and can be approximated by the Monte Carlo expression:

$$\frac{1}{N}\sum_{i=1}^{N} L_{t+1,T}\left(\theta^{(i)}\right) \tag{2.16}$$

where $\{\theta^{(i)} : i=1,N\}$ is a random sample of the normal distribution $N_{p(S)}(\hat{\theta}_{s,t}, B_{s,t}^{-1}(\hat{\theta}_{s,t}))$.

- *The reparameterised predictive density criterion (RPRED).*

Although the posterior distribution $\pi(\theta \mid \mathbf{F}_{s,t})$ is asymptotically normal, in small samples this approximation could be improved by taking into account that the parameters of the variance equation ($\omega, \alpha_1,...,\alpha_p, \beta_1,...., \beta_q$) are non-negative, and using the logaritmic parameterisation ($\log \omega, \log \alpha_1,..., \log \alpha_p, \log \beta_1,...., \log \beta_q$). In this case we could use the approximation:

$$\pi(g(\theta) \mid \mathbf{F}_{s,t}) \sim N_{p(S)}\left(g(\hat{\theta}_{s,t}), \left(\frac{dg}{d\theta}\left(\hat{\theta}_{s,t}\right)\right)' B_{s,t}^{-1}(\hat{\theta}_{s,t})\left(\frac{dg}{d\theta}\left(\hat{\theta}_{s,t}\right)\right)\right) \tag{2.17}$$

where $g(\theta) = (\mu, \gamma_1, ..., \gamma_{nar}, \theta_1, ..., \theta_{nma}, \log \omega, \log \alpha_1,..., \log \alpha_p, \log \beta_1,...., \log \beta_q)'$. The reparameterised predictive density criterion (RPRED) is based on expression (2.17) and is given by:

$$\mathrm{RPREDR}(Y_v \mid \mathbf{F}_{s,t},\mathbf{S}) =$$

$$= \int_{R^{p(S)}} L_{t+1,T}(g^{-1}(u)) N_{p(S)}\left(g(\hat{\theta}_{s,t}), \left(\frac{dg}{d\theta}\left(\hat{\theta}_{s,t}\right)\right)' B_{s,t}^{-1}(\hat{\theta}_{s,t})\left(\frac{dg}{d\theta}\left(\hat{\theta}_{s,t}\right)\right)\right)(u)du \tag{2.18}$$

This expression is also usually analitically intractable, and can be approximated by Monte Carlo methods in a similar way to the PRED criterion.

Asymptotics

This Section discusses the asymptotic behaviour of the information set selection criteria described above in ARMA-GARCH models. The results are stated in terms of the PICF criterion; however, the rest of the criteria have the same behaviour given that all of them are equivalent asymptotically [Muñoz et al. (2001)].

Theorem 1 analyses the ARMA-ARCH case, while Theorem 2 analyses the ARMA-GARCH case.

Theorem 1

Let $1\leq r<s<t$. Let $P_{o,s,t}$ be the restriction of P_o to the period $\{s,...,t\}$.

a) Let us assume that P_o is ARMA(nar,nma)-ARCH(p) with $\theta = \theta_o \in \text{int } \Theta$. If $E[\varepsilon_t^4(\theta_0)] < \infty$ and $t-s = o(\min\{t-r,T-t\})$ .Then:

$$\frac{\exp\left[\text{PICF}(Y_v \middle| \mathbf{F}_{s,t};\mathbf{S})\right]}{\exp\left[\text{PICF}(Y_v \middle| \mathbf{F}_{r,t};\mathbf{S})\right]} \to 0 \text{ with } P_{0,r,T}\text{-probability 1 when } t\text{-}s\to\infty \quad (2.19)$$

b) Let $R_1 = \{i\in\{r,....,T\}: P_{0,i,T} = \mathbf{S}(\theta_0)\}$ and $R_2 = \{r,....,T\}-R_1$, i.e., R_1 (resp. R_2) is the subperiod of $\{r,...,T\}$ where the proposed model is correctly specified (resp. incorrectly specified). If:

$$|R_2 \cap \{t+1,...,T\}| = o(|R_1 \cap \{t+1,...,T\}|)$$

$$|R_1 \cap \{r,...,s\}| = o(|R_2 \cap \{r,...,s\}|)$$

when $t\text{-}s\to\infty$, where $|A|$ denotes the cardinal of A, then:

$$\frac{\exp\left[\text{PICF}(Y_v \middle| \mathbf{F}_{s,t};\mathbf{S})\right]}{\exp\left[\text{PICF}(Y_v \middle| \mathbf{F}_{r,t};\mathbf{S})\right]} \to \infty \text{ with } P_{0,r,T}\text{-probability 1} \quad (2.20)$$

when $t\text{-}s\to\infty$.

Proof: See Appendix ◁

Theorem 2

Let us assume that P_0 belongs to the family **S** with $\theta = \theta_o \in \text{int } \Theta$ and further, that the conditional distribution D has a support with at least 3 elements and verifies that

$\eta_t = \dfrac{\varepsilon_t}{h_t^{1/2}}$ is an ergodic and strictly stationary process, uniformly bounded and $\alpha-$mixing of size $-\dfrac{r}{r-2}$, $r > 2$ or ϕ-mixing of size $-\dfrac{r}{2(r-1)}$, $r \geq 2$.

Under the conditions of Theorem 1 on T,t, s and r, (2.19) and (2.20) are satisfied with weak convergence. Furthermore, this convergence is strong if it is verified that $\left(\ln\left(\max\{t-r, T-t\}\right)\right)^{2/p(\mathbf{S})}$ (t-s) = o(min{t-r,T-t}) where p(**S**) is the number of parameters of the model **S**.

Proof: See Appendix ◃

Building Forecast Intervals

One of the most important objectives of the selection of the information set is to provide more accurate outsampling forecasts. In this section we show how to build k-steps approximated (1-α)-posterior Bayesian predictive intervals with $\alpha \in (0,1)$, using the asymptotic normality of the posterior distribution $\pi(\theta \mid \mathbf{F}_{s,t})$.

These intervals are based on the following approximation to the k-step joint posterior predictive distribution:

$$f(y_{T+1},\dots,y_{T+k} \mid \{y_s, \dots, y_T\}) = \int_\Theta f\left(y_{T+1},\dots,y_{T+k} \middle| \theta, y_s,\dots,y_T\right)\pi\left(\theta \middle| y_s,\dots,y_T\right)d\theta \approx$$

$$\approx \int_\Theta f\left(y_{T+1},\dots,y_{T+k} \middle| \theta, y_s,\dots,y_T\right) N_{p(S)}\left(\hat\theta_{s,T}, B_{s,T}^{-1}\left(\hat\theta_{s,T}\right)\right)d\theta \tag{2.21}$$

This expression is usually analitically intractable, but can be approximated by Monte Carlo methods. We can draw$\{\theta^{(i)}; i=1,..,N\}$ a random sample from the normal distribution $N_{p(S)}\left(\hat\theta_{s,T}, B_{s,T}^{-1}(\hat\theta_{s,T})\right)$ and, next draw $\{(y_{T+1}^{(i)},\dots,y_{T+k}^{(i)}); i=1,\dots,N\}$ a random sample from the distribution $f(y_{T+1},\dots,y_{T+k} \mid \theta^{(i)}, \{y_s, \dots, y_T\}))$ using the recursive expressions:

$$y_\ell^{(i)} = \mu^{(i)} + \sum_{j=1}^{nar}\gamma_j^{(i)} y_{\ell-j}^{(i)} + \sum_{j=1}^{nma}\theta_j^{(i)}\varepsilon_{\ell-j}^{(i)} + \varepsilon_\ell^{(i)} \quad \ell = T+1,\dots,T+k$$

$$h_\ell^{(i)} = \omega^{(i)} + \sum_{j=1}^{p}\alpha_j^{(i)}\varepsilon_{\ell-j}^{(i)2} + \sum_{j=1}^{q}\beta_j^{(i)} h_{\ell-j}^{(i)} \quad \ell = T+1,\dots,T+k \tag{2.22}$$

where

$$\theta^{(i)} = (\mu^{(i)}, \gamma_1^{(i)},\ldots,\gamma_{nar}^{(i)}, \theta_1^{(i)},\ldots,\theta_{nma}^{(i)}, \varpi^{(i)}, \alpha_1^{(i)},\ldots,\alpha_p^{(i)}, \beta_1^{(i)},\ldots,\beta_q^{(i)})'$$

$$\varepsilon_\ell^{(i)} \text{ is drawn from } N(0,\ h_\ell^{(i)}) \text{ if } \ell = T+1,\ldots,T+k$$

$$\varepsilon_\ell^{(i)} = y_1 - \mu^{(i)} - \sum_{j=1}^{nar}\gamma_j^{(i)} y_{\ell-j} - \sum_{j=1}^{nma}\theta_j^{(i)}\varepsilon_{\ell-j}^{(i)} \quad \text{if } \ell = 1,\ldots,T$$

$$\varepsilon_\ell^{(i)} = 0 \text{ if } \ell \le 0 \text{ ; } h_0^{(i)} = 0 \qquad (2.23)$$

The approximated (1-α)-posterior Bayesian predictive interval for y_{T+j} ; j=1,...,k will be given by :

$$\left[y_{T+j}\left(\frac{\alpha}{2}\right), y_{T+j}\left(1-\frac{\alpha}{2}\right)\right] \qquad (2.24)$$

where $y_{T+j}(\beta)$ denotes the β-quantile of the Monte Carlo sample $\left\{ y_{T+j}^{(1)},\ldots, y_{T+j}^{(N)} \right\}$.

3. Monte Carlo Study

In this Section, we carry out a simulation study using the above criteria in ARMA-GARCH models with our aim being to test the behaviour of these criteria in finite samples. We understand a criterion to be better when the size of the information set it selects is larger in the circumstances where there has been no structural change that justifies the selection of samller sets. This is due to the fact that when the size of the selected information set is larger, the size of the sample used in the estimation of the parameters of the model will be larger. Similarly, the precision with which these parameters are estimated will be greater and the accuracy of the forecast intervals that are built will also be greater.

Selecting the information set in ARMA-ARCH models

We have simulated 20 ARMA-ARCH models which are particular cases of the ARMA(1,1)-ARCH(2) given by the equations:

$$y_t = \gamma_1 y_{t-1} + \theta_1\varepsilon_{t-1} + \varepsilon_t \quad t=1,\ldots,T \qquad (3.1)$$

with $\varepsilon_t \mid \{y_1,\ldots,y_{t-1}\} \sim N(0,h_t)$ where:

$$h_t = \omega + \alpha_1\,\varepsilon_{t-1}^2 + \alpha_2\,\varepsilon_{t-2}^2 \ \text{ with } \ \varepsilon_t = 0 \ \text{ if } t \le 0 \qquad (3.2)$$

The values of the parameters of the simulated models are shown in Table 1. All the models have an inconditional variance equal to 1 and correspond to several ways of modeling the conditional mean equation with AR and MA effects of different intensity and several degrees of heteroscedasticity, persistence and parsimony in the conditional variance equation.

TABLE 1. Simulated ARMA-ARCH models

Model	**Equation**	γ_1	θ_1	ω	α_1	α_2
1	Arma(0,0)-Arch(0)	0.00	0.00	1.00	0.00	0.00
2	Arma(0,0)-Arch(1)	0.00	0.00	0.50	0.50	0.00
3	Arma(0,0)-Arch(1)	0.00	0.00	0.10	0.90	0.00
4	Arma(0,0)-Arch(2)	0.00	0.00	0.50	0.25	0.25
5	Arma(1,0)-Arch(0)	0.50	0.00	1.00	0.00	0.00
6	Arma(1,0)-Arch(1)	0.50	0.00	0.50	0.50	0.00
7	Arma(1,0)-Arch(1)	0.50	0.00	0.10	0.90	0.00
8	Arma(1,0)-Arch(2)	0.50	0.00	0.50	0.25	0.25
9	Arma(1,0)-Arch(0)	0.90	0.00	1.00	0.00	0.00
10	Arma(1,0)-Arch(1)	0.90	0.00	0.50	0.50	0.00
11	Arma(1,0)-Arch(1)	0.90	0.00	0.10	0.90	0.00
12	Arma(1,0)-Arch(2)	0.90	0.00	0.50	0.25	0.25
13	Arma(0,1)-Arch(0)	0.00	0.50	1.00	0.00	0.00
14	Arma(0,1)-Arch(1)	0.00	0.50	0.50	0.50	0.00
15	Arma(0,1)-Arch(1)	0.00	0.50	0.10	0.90	0.00
16	Arma(0,1)-Arch(2)	0.00	0.50	0.50	0.25	0.25
17	Arma(0,1)-Arch(0)	0.00	0.90	1.00	0.00	0.00
18	Arma(0,1)-Arch(1)	0.00	0.90	0.50	0.50	0.00
19	Arma(0,1)-Arch(1)	0.00	0.90	0.10	0.90	0.00
20	Arma(0,1)-Arch(2)	0.00	0.90	0.50	0.25	0.25

The size of the simulated periods has been T = 100 and 200 while the number of information set considered has been 3, with sizes 25 (Inf1) , 0.5T (Inf2) and T (Inf3). The number of replications has been 500 and the number of simulations to calculate the PRED and RPRED criteria has been 100.

Table 2 shows the proportion of times that each information set has been chosen. Note that for all the criteria the larger the value of T, the larger the information set that is chosen. The best peformance corresponds to the BICF, PRED and RPRED criteria, and there is no significant difference between them. The worst behaviour corresponds to the PICF in the less parsimonious models because, if the size of the information set is small, the determinant of the information matrix may be close to 0 if the likelihood function is flat. In that case, the penalization term will tend to favour the selection of smaller information sets.

TABLE 2. Results for ARMA-ARCH models

			BICF			PICF			PRED			RPRED	
Model	**T**	**Inf1**	**Inf2**	**Inf3**	**Inf1**	**Inf2**	**Inf3**	**Inf1**	**Inf2**	**Inf3**	**Inf1**	**Inf2**	**Inf3**
1	100	7.8	17.8	74.4	6.4	19.8	73.8	7.4	23.6	69.0	6.8	23.0	70.2
1	200	4.4	20.2	75.4	3.6	18.6	77.8	2.6	20.2	77.2	3.8	20.4	75.8
2	100	6.8	18.8	74.4	22.6	17.8	59.6	5.4	22.8	71.8	8.8	20.4	70.8
2	200	3.8	17.4	78.8	18.6	15.0	66.4	1.8	24.0	74.2	2.4	20.2	77.4
3	100	7.8	22.6	69.6	21.6	17.8	60.6	9.4	17.8	72.8	10.8	16.2	73.0
3	200	5.4	18.2	76.4	15.4	19.6	65.0	2.4	24.2	73.4	1.8	23.4	74.8
4	100	6.6	15.0	78.4	42.8	24.8	32.4	3.2	16.8	80.0	6.6	19.6	73.8
4	200	2.0	16.6	81.4	41.2	14.4	44.4	0.8	21.6	77.6	2.4	20.6	77.0
5	100	6.0	17.0	77.0	6.2	16.2	77.6	6.8	18.8	74.4	7.4	18.2	74.4
5	200	3.4	22.4	74.2	2.8	20.6	76.6	2.4	23.8	73.8	2.6	23.0	74.4
6	100	6.2	18.6	75.2	25.8	19.2	55.0	5.4	20.0	74.6	6.4	19.4	74.2
6	200	1.4	21.4	77.2	20.6	17.4	62.0	1.6	20.0	78.4	2.2	18.8	79.0
7	100	5.2	23.8	71.0	20.4	20.0	59.6	5.0	21.8	73.2	5.8	21.2	73.0
7	200	3.6	21.8	74.6	14.8	15.6	69.6	2.4	21.8	75.8	3.2	21.4	75.4
8	100	5.0	15.8	79.2	40.6	26.2	33.2	4.6	20.6	74.8	10.0	21.2	68.8
8	200	1.2	19.6	79.2	37.0	16.4	46.6	1.0	18.6	80.4	1.4	19.8	78.8
9	100	8.6	22.2	69.2	4.2	21.0	74.8	6.4	19.8	73.8	5.2	25.0	69.8
9	200	2.4	22.4	75.2	2.2	20.6	77.2	2.8	20.6	76.6	2.8	20.8	76.4
10	100	6.0	16.8	77.2	14.6	21.4	64.0	6.0	18.6	75.4	5.2	20.0	74.8
10	200	4.0	21.6	74.4	11.8	18.4	69.8	2.0	22.8	75.2	2.2	19.6	78.2
11	100	7.8	19.6	72.6	15.2	16.4	68.4	4.6	16.8	78.6	4.6	15.4	80.0
11	200	2.8	25.8	71.4	9.6	20.2	70.2	1.6	20.6	77.8	2.8	21.2	76.0
12	100	5.2	18.8	76.0	36	23.4	40.6	4.6	18.2	77.2	6.6	19.0	74.4
12	200	2.0	20.6	77.4	30	15.8	54.2	0.6	18.8	80.6	1.0	16.6	82.4
13	100	5.2	21.2	73.6	13.4	18.4	68.2	6.2	20.0	73.8	6.2	22.2	71.6
13	200	3.0	20.4	76.6	6.8	24.2	69.0	3.0	25.6	71.4	3.0	23.6	73.4
14	100	8.0	19.8	72.2	28.2	18.4	53.4	4.4	19.0	76.6	5.8	19.0	75.2
14	200	3.0	22.0	75.0	21.4	14.8	63.8	1.2	17.8	81.0	2.2	19.4	78.4
15	100	12.0	19.8	68.2	23.2	22.0	54.8	5.8	22.6	71.6	5.2	23.6	71.2
15	200	7.8	21.8	70.4	22.6	20.4	57.0	4.0	23.6	72.4	3.8	26.2	70.0
16	100	5.4	18.4	76.2	42.2	22.6	35.0	3.8	14.6	81.6	7.0	16.6	76.4
16	200	3.6	18.8	77.6	40.6	11.4	48.0	0.6	18.6	80.8	2.6	21.6	75.8
17	100	15.2	25.2	59.6	8.2	26.6	65.2	8.0	25.8	66.2	8.6	24.6	66.8
17	200	4.6	23.2	72.2	4.0	24.4	71.6	2.6	25.0	72.4	3.0	27.4	69.6
18	100	13.4	24.4	62.2	23.4	22.0	54.6	9.2	20.6	70.2	10.4	22.2	67.4
18	200	6.0	22.2	71.8	14.6	21.8	63.6	1.0	23.2	75.8	2.2	23.0	74.8
19	100	18.0	27.8	54.2	23.0	32.2	44.8	6.4	25.2	68.4	8.4	26.6	65.0
19	200	13.8	31.0	55.2	19.0	28.8	52.2	4.6	28.2	67.2	4.0	29.2	66.8
20	100	14.8	20.4	64.8	26.8	32.0	41.2	5.6	19.8	74.6	8.4	21.6	70.0
20	200	4.6	21.6	73.8	26.6	23.0	50.4	0.2	22.6	77.2	1.6	19.6	78.8

Selecting the information set in ARMA-GARCH models

Table 3 presents the 6 simulated models, which are particular cases of the ARMA(0,0)-GARCH(1,1) model:

$$y_t = \varepsilon_t \quad t=1,...,T \tag{3.3}$$

with $\varepsilon_t \mid \{y_1,...,y_{t-1}\} \sim NT_k(0,h_t)$ where:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \text{ where } h_t = 0 \text{ and } \varepsilon_t = 0 \text{ if } t \le 0 \tag{3.4}$$

where $NT_k(0,\sigma^2)$ denotes the [-k,k]-truncated normal distribution. The simulated models correspond to several combinations of persistence and intensity of the ARCH effect and all the models have an inconditional variance equal to 1. In all cases k=10.

TABLE 3. Simulated ARMA(0,0)-GARCH(1,1) models

Model	ω	α_1	β_1
1	0.5	0.1	0.4
2	0.5	0.25	0.25
3	0.5	0.4	0.1
4	0.1	0.1	0.8
5	0.1	0.45	0.45
6	0.1	0.8	0.1

The simulated period and the considered information set have been the same as in Section 3.1. The number of replications has been 500 and the number of simulations to calculate the PRED and RPRED criteria has been 100.

Table 4 shows the proportions of times that each information set has been chosen. Note that the observed behaviour of the criteria is similar to the ARMA-ARCH case.

TABLE 4. Results for ARMA(0,0)-GARCH models

			BICF			PICF			PRED			RPRED	
Model	**T**	**Inf1**	**Inf2**	**Inf3**	**Inf1**	**Inf2**	**Inf3**	**Inf1**	**Inf2**	**Inf3**	**Inf1**	**Inf2**	**Inf3**
1	100	7.2	20.8	72.0	32.0	28.2	39.8	11.8	20.8	67.4	12.2	22.4	65.4
1	200	3.6	20.8	75.6	31.6	29.0	39.4	4.0	22.6	73.4	6.8	27.2	66.0
2	100	7.0	20.6	72.4	39.8	27.4	32.8	9.2	17.2	73.6	8.8	21.6	69.6
2	200	1.8	22.4	75.8	37.8	28.2	34.0	2.2	20.6	77.2	4.8	19.2	76.0
3	100	7.2	16.8	76.0	37.2	28.0	34.8	7.8	19.8	72.4	11.8	19.0	69.2
3	200	0.8	22.8	76.4	41.2	18.2	40.6	2.6	18.6	78.8	3.6	21.6	74.8
4	100	8.4	19.0	72.6	31.8	22.6	45.6	5.6	20.6	73.8	4.4	21.8	73.8
4	200	3.0	24.2	72.8	29.2	19.0	51.8	2.8	21.2	76.0	2.4	22.2	75.4
5	100	3.4	21.4	75.2	33.0	20.0	47.0	6.6	17.8	75.6	6.2	16.4	77.4
5	200	2.4	21.4	76.2	26.0	16.6	57.4	2.2	21.8	76.0	1.0	18.6	80.4
6	100	7.8	20.0	72.2	26.0	35.6	38.4	10.8	19.0	70.2	12.0	21.4	66.6
6	200	2.2	22.0	75.8	28.6	28.4	43.0	5.0	27.8	67.2	4.0	24.4	71.6

In summary, the performance of the BICF, PRED and RPRED criteria is similar. The worst performance corresponds to the PICF criterion when the DGP corresponds to less parsimonious models. This is the case because, when the sample size is small, the likelihood function becomes flat and, consequently, the determinant of the information matrix is close to cero and the penalisation term of this criterion favours the selection of smaller information sets.

4. Conclusion

In this paper, the performance of 4 information set selection criteria proposed in Muñoz et al. (2001) has been analysed in ARMA-GARCH models. We have found that the BICF, PRED and RPRED criteria show the best performance. By contrast, the PICF criterion of Phillips (1996) has a worse performance when the DGP is less parsimonious and the size of the information set is small. This occurs, because the likelihood may be flat and, as a consequence, the penalisation term of this criterion tends to favour the choice of smaller information sets.

Whe have also proposed an algorithm to build approximated Bayesian predictive intervals that takes into account the uncertainty associated with the estimation of the parameters of the model.

As regards future lines of research, we would recall that in this paper it has been assumed that the form of the model of the DGP has not changed overtime. Thus, it would be interesting to weaken this hipothesis by extending the proposed methodology to situations where there are several distinct families availables to explain the evolution of the series being considered.

Appendix

Proof of Theorem 1

The proof is based on Proposition 4.2 of Sin and White (1996) and theorem 4.6 of Muñoz et al. (2001). We start by checking that the conditions of this Proposition are satisfied.

a) Is trivial to check that $\ell_T(\theta)$ verifies conditions A(i), (ii) and (iii) of Sin and White (1996).

b) Let $\overline{\ell}(\theta)$ = -0.5E[log h_t]-0.5E[$\varepsilon_t^2 h_t^{-1}$]. Applying Theorem 3.1 of Weiss(1986) it follows that $\exists$ E[$\ell_T(\theta)$] = -0.5TE[log h_t]-0.5TE[$\varepsilon_t^2 h_t^{-1}$] = T$\overline{\ell}(\theta)$ a.s (P_0). Furthermore, $\overline{\ell}(\theta)$ is continuous on θ because log h_t + $\varepsilon_t^2 h_t^{-1}$ is continuous, Θ is compact and applying Theorem 9.30 of Davidson (1994). Therefore, condition A(iv) of Sin and White (1996) is satisfied.

c) E[$\ell_T(\theta)$] is continuously differentiable on θ and $E\left[\frac{d\ell_T}{d\theta}(\theta)\right] = \frac{d}{d\theta}E[\ell_T(\theta)]$, applying Theorem 9.31 of Davidson (1994). Note that $\ell_T(\theta)$ is continuously differentiable on θ, Θ is compact and $E\left[\frac{d\ell_T}{d\theta}(\theta)\right] < \infty$. This last statement is shown because $\frac{d\ell_T}{d\theta} = \frac{1}{2}\sum_{t=nar+1}^{T}\left(\left(1-\frac{\varepsilon_t^2}{h_t}\right)\frac{1}{h_t}\frac{dh_t}{d\theta}+\frac{2\varepsilon_t}{h_t}\frac{d\varepsilon_t}{d\theta}\right)$ and by triangular and Hölder inequalities :

$$E\left[\left\|\left(\left(1-\frac{\varepsilon_t^2}{h_t^2}\right)\frac{1}{h_t}\frac{dh_t}{d\theta}+\frac{2\varepsilon_t}{h_t}\frac{d\varepsilon_t}{d\theta}\right)\right\|\right] \leq$$

$$\leq E\left[\left\|\left(1-\frac{\varepsilon_t^2}{h_t^2}\right)\right\|^2\right]^{1/2} E\left[\frac{1}{h_t^2}\left\|\frac{dh_t}{d\theta}\right\|^2\right]^{1/2} + 2E\left[\left|\frac{\varepsilon_t}{h_t}\right|^2\right]^{1/2} E\left[\left\|\frac{d\varepsilon_t}{d\theta}\right\|^2\right]^{1/2} < \infty \qquad \text{(A.1)}$$

where (A.1) is shown taking into account that $h_t \geq \omega$ and that by recursive expressions (2.11), (2.12) and using that $\Theta \subseteq A$ given in (2.3), it follows that $\frac{d\varepsilon_t}{d\theta}$ =

$g_t(A_1,...,A_t)$ and $\frac{dh_t}{d\theta} = g_{1t}(U_1,...,U_t)$ where $U_t = D_t + 2\sum_{i=1}^{p} \alpha_i \varepsilon_{t-i} \frac{d\varepsilon_{t-i}}{d\theta}$ and g_t are g_{1t} polynomial functions in their arguments with exponentially bound coefficients. Therefore, $\exists M_1$ and $M_2 < \infty$ such that $E\left[\left\|\frac{d\varepsilon_t}{d\theta}\right\|^2\right] < M_1$ and $E\left[\frac{1}{h_t^2}\left\|\frac{dh_t}{d\theta}\right\|^2\right] < M_2$ because $E\left[\varepsilon_t^4\right] < \infty$. This follows because, applying the Mean Value Theorem, $\varepsilon_t(\theta) = \varepsilon_t(\theta_0) + \frac{d\varepsilon_t}{d\theta'}(\theta^*)(\theta - \theta_0)$ with $\theta^* \in [\theta,\theta_o]$, Θ is compact and $E[\varepsilon_t^4(\theta_0)] < \infty$ by hypothesis.

Therefore, condition A(v) of Sin and White (1996) is satisfied.

d) $\overline{\ell}(\theta)$ has a unique maximum in $\theta = \theta_o$ (Corollary 3.1 of Weiss (1986)) and $\theta_o \in \mathrm{int}\,\Theta$ by assumption. Therefore, conditions A(vi) and A(vii) of Sin and White(1996) are satisfied because $E[\ell_T(\theta)]$ is continuous on θ.

e) $u_t(\theta) = -\frac{1}{2}\log(h_t) - \frac{\varepsilon_t^2}{2h_t} \in C^2$ and, therefore, condition (i) of Proposition 4.1 of Sin and White (1996) is satisfied.

f) $\overline{\ell}_T(\theta) = \frac{\ell_T(\theta)}{T} \to \overline{\ell}(\theta)$ a.s (P_0) by Corollary 3.1 of Weiss (1986)

Furthermore, this convergence is uniform on Θ because

$$\forall \theta,\theta' \in \Theta \quad |u_t(\theta) - u_t(\theta')| \le \frac{1}{2}\left(|\log(h_t(\theta) - \log(h_t(\theta'))| + \left|\frac{\varepsilon_t^2(\theta)}{h_t(\theta)} - \frac{\varepsilon_t^2(\theta')}{h_t(\theta')}\right|\right)$$

If $\|.\|$ denotes the L_1-norm and applying the Theorem of Mean Value, it follows that:

$$|\log(h_t(\theta) - \log(h_t(\theta'))| \le \left\|\frac{dh_t}{d\theta}(\theta^*)h_t^{-1}(\theta^*)\right\| \, \|\theta - \theta'\|$$

where $\theta^* \in [\theta,\theta']$. Now, by Lemmas 3.1 and 3.2 of Weiss (1986), the Hölder inequality and taking into account that Θ is compact and that $h_t(\theta) \ge \omega \ \forall\theta$, it follows that:

$$E\left[\left\|\frac{dh_t}{d\theta}(\theta^*)h_t^{-1}(\theta^*)\right\|\right] \le \left(E\left[\frac{1}{h_t^2(\theta^*)}\frac{dh_t}{d\theta'}(\theta^*)\frac{dh_t}{d\theta}(\theta^*)\right]\right)^{1/2} < \infty$$

Similarly, it can be shown that:

$$\left|\frac{\varepsilon_t^2(\theta)}{h_t(\theta)} - \frac{\varepsilon_t^2(\theta')}{h_t(\theta')}\right| \leq \left\| 2\frac{\varepsilon_t(\theta^*)}{h_t(\theta^*)}\frac{d\varepsilon_t}{d\theta}(\theta^*) + \frac{\varepsilon_t^2(\theta^*)}{h_t^2(\theta^*)}\frac{dh_t}{d\theta}(\theta^*)\right\| \, \|\theta - \theta'\|$$

However,

$$E\left[\left\| 2\frac{\varepsilon_t(\theta^*)}{h_t(\theta^*)}\frac{d\varepsilon_t}{d\theta}(\theta^*) + \frac{\varepsilon_t^2(\theta^*)}{h_t^2(\theta^*)}\frac{dh_t}{d\theta}(\theta^*)\right\|\right] \leq$$

$$\leq E\left[\left\| 2\frac{\varepsilon_t(\theta^*)}{h_t(\theta^*)}\frac{d\varepsilon_t}{d\theta}(\theta^*)\right\|\right] + E\left[\left\|\frac{\varepsilon_t^2(\theta^*)}{h_t^2(\theta^*)}\frac{dh_t}{d\theta}(\theta^*)\right\|\right] \leq$$

$$\leq 2\left(E\left[\frac{\varepsilon_t^2(\theta^*)}{h_t(\theta^*)}\right]\right)^{1/2}\left(E\left[\frac{1}{h_t(\theta^*)}\frac{d\varepsilon_t}{d\theta'}(\theta^*)\frac{d\varepsilon_t}{d\theta}(\theta^*)\right]\right)^{1/2} +$$

$$+\left(E\left[\frac{\varepsilon_t^4(\theta^*)}{h_t^2(\theta^*)}\right]\right)^{1/2}\left(E\left[\frac{1}{h_t^2(\theta^*)}\frac{dh_t}{d\theta'}(\theta^*)\frac{dh_t}{d\theta}(\theta^*)\right]\right)^{1/2} < \infty$$

by Lemmas 3.1 and 3.2 of Weiss (1996) and taking into account that Θ is compact, $h_t(\theta) \geq \omega$ and $E\left[\varepsilon_t^4(\theta^*)\right] < \infty$.

By corollary 2 of Andrews (1987), it follows that the convergence of f) is uniform on Θ . Therefore, condition (ii) of Proposition 4.1 of Sin and White (1996) is satisfied.

g) Using a proof similar to that of Theorem 3.3 of Weiss (1986), it can be shown that each element of $\frac{du_t}{d\theta}(\theta_0)$ satisfies a Limit Central Theorem and, therefore, condition (iii) of Proposition 4.1 of Sin and White (1996) is satisfied.

h) Each element of $\frac{d^2u_t}{d\theta d\theta'}$ satisfies a UWLLN because, by Lemma 3.3 of Weiss (1986), it follows that $\exists\ E\left[\frac{d^2u_t}{d\theta d\theta'}\right]$ and, by the ergodic theorem $\frac{1}{T}\sum_{t=1}^{T}\frac{d^2u_t}{d\theta d\theta'} \to E\left[\frac{d^2u_t}{d\theta d\theta'}\right]$ with P_o probability 1. Furthermore, this convergence is uniform by corollary 2 of Andrews (1987) and following a similar proof to f) applied to each element of $\frac{d^2u_t}{d\theta d\theta'}$. Note that the moment of the derivatives of these elements are bounded because of the demonstration of Theorem 3.2 of Weiss (1986), page 129 paragraph (iii). Therefore condition (iv) (γ) of Proposition 4.1 of Sin and White (1996) is satisfied.

Condition (iv) (β) of Proposition 4.1 of Sin and White (1996) follows beacuse $E\left[\frac{d^2u_t}{d\theta d\theta'}\right]<\infty$. Finally, condition (iv) (α) is satisfied because $\frac{1}{T}\sum_{t=1}^{T}\frac{d^2u_t}{d\theta d\theta'} \to E\left[\frac{d^2u_t}{d\theta d\theta'}\right]$ and using that by Lemma 3.3 of Weiss (1986), $E\left[\frac{d^2u_t}{d\theta d\theta'}\right]$ is a positive definite matrix.

Therefore, the conditions of Proposition 4.2 of Sin and White (1996) are satisfied. Furthermore:

$$\frac{1}{t-s}B_{s,t}\left(\hat{\theta}_{s,t}\right)\to$$

$$0.5E\left[h_t^{-2}(\theta_0)\frac{dh_t}{d\theta}(\theta_0)\left(\frac{dh_t}{d\theta}(\theta_0)\right)'\right]+E\left[h_t^{-1}(\theta_0)\frac{d\varepsilon_t}{d\theta}(\theta_0)\left(\frac{d\varepsilon_t}{d\theta}(\theta_0)\right)'\right]$$

with P_o-probability 1 if t-s $\to \infty$ by Theorem 3.3 of Weiss (1986)) and, therefore, $B_{s,t}\left(\hat{\theta}_{s,t}\right) = O_p(t\text{-}s)$.◁

Proof of Theorem 2

Note that the series {y_t, h_t, ε_t, t=1,2,...} are uniformly bounded because of the uniform boundary condition on η_t, using the recursive equations (2.11) and (2.12), the compacity of Θ and the assumptions made on the root of the polinomials of the set A defined in (2.3). Therefore, the vectors A_t and D_t defined in (2.13) and (2.14), respectively, and the series $\{\frac{d\varepsilon_t}{d\theta}, \frac{dh_t}{d\theta}; t=1,2,...\}$ are uniformly bounded. Furthermore, from (2.11) and (2.12) and derivating with respect θ, it follows that :

$$\frac{d^2h_t}{d\theta d\theta'} = 2\sum_{k=1}^{p}\alpha_k \frac{d\varepsilon_{t-k}}{d\theta}\frac{d\varepsilon_{t-k}}{d\theta'} + 2\sum_{k=1}^{p}\alpha_k\varepsilon_{t-k}\frac{d^2\varepsilon_{t-k}}{d\theta d\theta'} + 2\sum_{k=1}^{p}\varepsilon_{t-k}\frac{d\varepsilon_{t-k}}{d\theta}\frac{d\alpha_k}{d\theta'} +$$

$$+\sum_{k=1}^{q}\beta_l\frac{d^2h_{t-k}}{d\theta d\theta'} + \sum_{l=1}^{q}\frac{dh_{t-k}}{d\theta}\frac{d\beta_k}{d\theta'} + \frac{dB_t}{d\theta'} \tag{A.2}$$

$$\frac{d^2\varepsilon_t}{d\theta d\theta'} = \frac{dA_t}{d\theta'} - \sum_{j=1}^{nma}\frac{d\varepsilon_{t-j}}{d\theta}\frac{d\theta_j}{d\theta'} - \sum_{j=1}^{nma}\theta_j\frac{d^2\varepsilon_{t-j}}{d\theta d\theta'} \tag{A.3}$$

and following a similar reasoning to $\frac{d\varepsilon_t}{d\theta}$ and $\frac{dh_t}{d\theta}$ it can be shown that the series $\{\frac{d^2\varepsilon_t}{d\boldsymbol{\theta} d\boldsymbol{\theta}'}, \frac{d^2h_t}{d\theta d\theta'}; t=1,2,...\}$ are uniformly bounded.

The proof of the theorem is based on checking that the conditions of example 7.1 of Sin and White (1996) are satisfied..

A(i) $\forall\theta\in\Theta$ $\ell_T(\theta)$ is {y_1,...,y_T}-measurable because the series { $u_t(\theta)$ t=1,...,T} are {y_1,...,y_T}-measurable

A(ii) $\ell_T(\theta)$ is continuous on Θ because $\varepsilon_t(\theta)$ and $h_t(\theta)$ are continuous on θ $\forall$t=1,...,T

A(iii) $\ell_T(\theta)$ is continuously differentiable on θ because $\varepsilon_t(\theta)$ and $h_t(\theta)$ are continuous differentiable on θ $\forall$t=1,...,T

A(iv) $\forall\theta\in\Theta$ $\exists$ $E[\ell_T(\theta)]$ that is a continuous function on θ.

It is only necessary to check that $\exists$ $E[u_T(\theta)]$ is continuous on θ. However:

$$E\left[\left|u_t(\theta)\right|\right] \le \frac{1}{2}\left(E\left[\left|\log \ell_T(\theta)\right|\right] + E\left[\frac{\varepsilon_t^2(\theta)}{2h_t(\theta)}\right]\right) < \infty$$

because $h_t(\theta)$ and $\varepsilon_t^2(\theta)$ are uniformly bounded and $h_t(\theta) \ge \omega \quad \forall \theta \in \Theta$, t=1,...,T

The continuity of $E[u_t(\theta)]$ is shown by Theorem 9.30 of Davidson (1994).

A(v) $E[\ell_T(\theta)]$ is continuously differentiable on Θ and $\frac{d}{d\theta}E[\ell_T(\theta)] = E\left[\frac{d}{d\theta}\ell_T(\theta)\right]$

This can be shown using the uniform boundary of $\{u_t(\theta)$;t=1,2,...$\}$ and applying Theorem 9.31 of Davidson (1994)

A(vi) If $\theta_T^* = \arg\max_{\theta\in\Theta} \frac{1}{T}E[\ell_T(\theta)]$ then $\theta_T^* = \theta_o$ (θ_T^* exists because Θ is compact and $E[\ell_T(\theta)]$ is continuous on θ by A(iv)).

In order to check this condition we will follow a similar reasoning to Theorem 3.1 of Weiss (1986) or Theorem 1 of Lumsdaine (1996).

Because of the uniform boundary of η_t and the stationarity of y_t, it follows that :

$$\frac{1}{T}E[\ell_T(\theta)] = E[u_t(\theta)] \tag{A.4}$$

However,

$$2\left(E[u_t(\theta)] - E[u_t(\theta_o)]\right) = -E\left[\log\frac{h_t(\theta)}{h_t(\theta_o)}\right] - E\left[\frac{\varepsilon_t^2(\theta)}{h_t(\theta)} - \frac{\varepsilon_t^2(\theta_o)}{h_t(\theta_o)}\right] \tag{A.5}$$

Furthermore:

$$E\left[\frac{\varepsilon_t^2(\theta)}{h_t(\theta)} - \frac{\varepsilon_t^2(\theta_o)}{h_t(\theta_o)}\right] = E\left[\frac{(\varepsilon_t(\theta) - \varepsilon_t(\theta_o) + \varepsilon_t(\theta_o))^2}{h_t(\theta)} - \frac{\varepsilon_t^2(\theta_o)}{h_t(\theta_o)}\right] =$$

$$= E\left[\frac{(\varepsilon_t(\theta) - \varepsilon_t(\theta_o))^2}{h_t(\theta)}\right] + 2E\left[\varepsilon_t(\theta_o)\frac{\varepsilon_t(\theta) - \varepsilon_t(\theta_o)}{h_t(\theta)}\right] + E\left[\frac{\varepsilon_t^2(\theta_o)}{h_t(\theta)}\right] - E\left[\frac{\varepsilon_t^2(\theta_o)}{h_t(\theta_o)}\right] =$$

$$\left(E\left[\varepsilon_t(\theta_o)\frac{\varepsilon_t(\theta) - \varepsilon_t(\theta_o)}{h_t(\theta)}\right] = 0 \text{ because } E\left[\varepsilon_t(\theta_o) \middle| \Phi_{t-1}\right] = 0 \text{ and } \frac{\varepsilon_t(\theta) - \varepsilon_t(\theta_o)}{h_t(\theta)} \text{ function of } \Phi_{t-1} \right)$$

$$= E\left[\frac{(\varepsilon_t(\theta)-\varepsilon_t(\theta_o))^2}{h_t(\theta)}\right] + E\left[\frac{h_t(\theta_o)}{h_t(\theta)}\right] - 1 \tag{A.6}$$

From (A.5) and (A.6) it follows that :

$$2\left(E[u_t(\theta)] - E[u_t(\theta_o)]\right) = E\left[\frac{-h_t(\theta_o)}{h_t(\theta)} + \log\frac{h_t(\theta_o)}{h_t(\theta)}\right] - E\left[\frac{(\varepsilon_t(\theta)-\varepsilon_t(\theta_o))^2}{h_t(\theta)}\right] + 1 \le$$

$$\le -E\left[\frac{(\varepsilon_t(\theta)-\varepsilon_t(\theta_o))^2}{h_t(\theta)}\right] \le 0 \tag{A.7}$$

because the function $f(x) = -\ln x + x \ge 0$ and it has an absolute maximum on x=1 with f(1)=1. Then, from (A.4):

$$\forall\theta\in\Theta \quad \frac{1}{T}E[\ell_T(\theta)] \le \frac{1}{T}E[\ell_T(\theta_0)] \tag{A.8}$$

with equality if and only if $\varepsilon_t(\theta) = \varepsilon_t(\theta_o)$ and, therefore, the parameters of the mean equation of the model will be equal due to the identificability of the ARMA process. Thus, if we divide θ in accordance with the conditional mean and variance equations such that $\theta = \left(\theta_1'.\theta_2'\right)'$ and $\theta_0 = \left(\theta_{01}'.\theta_{02}'\right)'$, it follows that $\theta_1 = \theta_{01}$.

On the other hand $h_t(\theta) = h_t(\theta_o)$ implies that :

$$\exists\, E\left[\left(\frac{dh_t}{d\theta_2}\right)\left(\frac{dh_t}{d\theta_2}\right)'\right] \text{ and } \det E\left[\left(\frac{dh_t}{d\theta_2}\right)\left(\frac{dh_t}{d\theta_2}\right)'\right] > 0$$

because if $\exists\ \lambda \ne 0$ such that $\lambda'\frac{dh_t}{d\theta_2} = 0$ a.s. (P_0) and by the recursive equations (2.11) and (2.12) it would follow that :

$$\sum_{k=1}^{p}\alpha_k\varepsilon_{t-k}\lambda'\left(\frac{d\varepsilon_{t-k}}{d\theta_2}\right) + \lambda'V_t = 0 \tag{A.9}$$

where $V_t = (-1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-nma}^2, h_{t-1}, .., h_{t-q})'$. Taking into account that $h_{t-i} = f(\varepsilon_{t-i-1}^2, \ldots, \varepsilon_o^2)$ and that by the Theorem of the Mean Value

$\varepsilon_t(\theta)=\varepsilon_t(\theta_o)+\frac{d\varepsilon_t(\theta^*)}{d\theta'}(\theta-\theta_o)$ where $\theta^*\in[\theta,\theta_o]$, then (A.9) would be a quadratic form on $\varepsilon_t(\theta_o)$ that can be solved by $\varepsilon_t(\theta_o)=f_1(t)$ or $\varepsilon_t(\theta_o)=f_2(t)$. However, by assumption, the support of the distribution of $\varepsilon_t(\theta_o)$ has at least 3 distinct elements, in contradiction to (A.9). Therefore, $E\left[\left(\frac{dh_t}{d\theta_2}\right)\left(\frac{dh_t}{d\theta_2}\right)'\right]>0$. Furthermore, taking into account that the mean parameters are equal and applying the Theorem of the Mean Value $\exists\,\theta_2^*\in[\theta_2,\theta_{20}]$ such that:

$$h_t(\theta)=h_t(\theta_o)+\frac{dh_t}{d\theta_2'}(\theta_2^*)(\theta_2-\theta_{20})\Rightarrow$$

$$\Rightarrow E\left[h_t^2(\theta)\right]=E\left[h_t^2(\theta_o)\right]+(\theta_2-\theta_{2o})'E\left[\frac{dh_t}{d\theta_2}\left(\frac{dh_t}{d\theta_2}\right)'\right](\theta_2-\theta_{2o})\geq E\left[h_t^2(\theta_o)\right]$$

with equality if and only if $\theta_2=\theta_{20}$. In summary, that implies $\theta=\theta_o$. From (A.8) it follows that $\theta_T^*=\theta_0$. Furthermore, by assumption $\theta_0\in\operatorname{int}\Theta$ and then A(vi) is satisfied.

A(vii) (Identifiable Uniqueness), i.e., given $\varepsilon>0$ $\exists N_0(\varepsilon)<\infty$ and $\delta(\varepsilon)>0$ such that:

$$\inf\left\{\min K_T^*(\theta):\theta\in N_T^*(\varepsilon)^c, T>N_0(\varepsilon)\right\}\equiv\delta(\varepsilon)$$

where $K_T^*(\theta)=\frac{1}{T}E\left[\ell_T(\theta_T^*)\right]-\frac{1}{T}E\left[\ell_T(\theta)\right]$ and $N_T^*(\varepsilon)^c$ is the compact complement of $N_T^*(\varepsilon)=B_T^*(\varepsilon)\bigcap\Theta$ where $B_T^*(\varepsilon)$ is an open sphere centred at θ_T^* with fixed radius ε.

However, from (A.8) it follows that :

$$\forall\,\theta\in N_T^*(\varepsilon)^c\quad K_T^*(\theta)=\frac{1}{T}\left(E\left[\ell_T(\theta_o)\right]-\frac{1}{T}E\left[\ell_T(\theta)\right]\right)>0$$

Taking into account that $K_T^*(\theta)$ is continuous on θ and that $N_T^*(\varepsilon)^c$ is compact, A(vii) it is proved.

Next we check that conditions (i) to (iv) of example 7.1 of Sin and White (1996) are satisfied.

1)

(α) $\{y_t\}$ is NED (Near Epoch Dependent Process) on $\{\eta_t\}$ of size -1/2.

This follows by applying Proposition AII.1 of Sin and White (1992) where conditions (ii) and (iii) of this Proposition are satisfied due to the uniform boundary of ε_t, h_t and y_t and conditions (i) and (iv) are satisfied because the DGP is an ARMA(nar,nma) process where the roots of the autorregresive polynomial are outside the unit circle.

(β) $\forall$ t $\frac{1}{h_t(\theta_o)}, \frac{dh_t}{d\theta}(\theta_o), \varepsilon_t(\theta_o)$ and $\frac{d\varepsilon_t}{d\theta}(\theta_o)$ are uniformly bounded, as has been shown above.

(γ) Θ is such that $\sum_{i=1}^{nar}|\gamma_i| < 1$, $\sum_{i=1}^{nar}|\beta_i| < 1$ because of the definition of A in (2.3)

(ii) $u_t(\theta)$ satisfies an USLLN on Θ, i.e.:

a) $\forall\theta \in \Theta \frac{1}{T}\sum_{t=1}^{T} E[u_t(\theta)]$ exists and is continuous on θ uniformly in T.

That is true because of A (iv) and the ergodicity and strict stationarity of the process $\{u_t(\theta), t=1,...,T\}$, which implies that:

$$\frac{1}{T}\sum_{t=1}^{T} E[u_t(\theta)] = E[u_t(\theta)]$$

b) $$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}(u_t - E[u_t(\theta)])\right| = o(1) \quad \text{a.s. } (P_o) \tag{A.10}$$

In order to show (A.10) we will use Theorem 21.8 of Davidson (1994). By the ergodic theorem:

$$\frac{1}{T}\sum_{t=1}^{T}\left(u_t(\theta) - E[u_t(\theta)]\right) \xrightarrow{a.s.} 0 \quad \forall\theta \in \Theta \tag{A.11}$$

Furthermore :

$$\frac{1}{T}\sum_{t=1}^{T}\left(u_t(\theta) - E[u_t(\theta)]\right) \text{ is strongly equicontinuous on } \Theta \tag{A.12}$$

This condition will be shown using Theorem 21.11 of Davidson (1994).

$\forall\theta,\theta'\in\Theta$ and applying the Theorem of Mean Value and the Hölder inequality :

$$\frac{\left|u_t(\theta)-u_t(\theta')\right|}{T}=\frac{\left|\frac{du_t}{d\theta'}\left(\theta^*\right)(\theta-\theta')\right|}{T}\le\sup\frac{1}{T}\left\|\frac{du_t}{d\theta'}\right\|_2\left\|\theta-\theta'\right\|_2$$

where $\theta^*\in[\theta,\theta']$.

However, $\frac{du_t}{d\theta}$ is uniformly bounded and ergodic; thus:

$$\sum_{t=1}^{T}\frac{1}{T}\left\|\frac{du_t}{d\theta'}\right\|_2=\left\|\frac{du_t}{d\theta'}\right\|_2<\infty$$

Therefore, the assumptions of Theorem 21.11 of Davidson (1994) and (A.12) is proved. From (A.11), (A.12) and Theorem 21.8 of Davidson(1994) , (ii) is proved.

(iii)

(α) $\exists\ \varepsilon,\ \alpha>0$ such that if T is large enough and $\forall\,\theta\in N_T^*(\varepsilon)$ then :

$$\det\left(\frac{1}{T}\frac{d^2}{d\theta d\theta'}\ell_T(\theta)\right)\ge\alpha\quad\text{a.s. }(P_o)\tag{A.13}$$

This is true because by the ergodic theorem:

$$\frac{1}{T}\frac{d^2}{d\theta d\theta'}\ell_T(\theta)=\frac{1}{T}\sum_{t=na+1}^{T}\frac{d^2}{d\theta d\theta'}u_t(\theta)\longrightarrow E\left[\frac{d^2}{d\theta d\theta'}u_t(\theta)\right]\ \text{a.s. }(P_0)\tag{A.14}$$

However,

$$E\left[\frac{d^2}{d\theta d\theta'}u_t(\theta)\right]=E\left[-\frac{1}{2h_t^2}\frac{dh_t}{d\theta}\frac{dh_t}{d\theta'}-\frac{1}{h_t}\frac{d\varepsilon_t}{d\theta}\frac{d\varepsilon_t}{d\theta'}\right]<0\tag{A.15}$$

by a similar proof to A (vi) . Then, (A.12) is proved from (A.14) and (A.15).

(β) If T is large enough and $\theta\in N_T^*(\varepsilon)$ then:

$$E\left[\frac{1}{T}\frac{d^2}{d\theta d\theta'}\ell_T(\theta)\right]=E\left[\frac{d^2}{d\theta d\theta'}u_t(\theta)\right]\ \text{ is }O(1)\ \text{ a.s. }(P_0)$$

due to the uniform boundary of $\varepsilon_t^2, \frac{dh_t(\theta)}{d\theta}, \frac{d^2h_t(\theta)}{d\theta d\theta'}$ and $\frac{d^2\varepsilon_t(\theta)}{d\theta d\theta'}$ and that $h_t(\theta) \geq$ ω

(γ) Each element of $\frac{d^2u_t(\theta)}{d\theta d\theta'}$ satisfies a USLLN on θ

This can be shown following a similar proof to (ii), applying the ergodic theorem and taking into account that the elements of $\frac{d^2u_t(\theta)}{d\theta d\theta'}$ are continuous functions of the elements of $\frac{1}{h_t}, \varepsilon_t, \frac{dh_t}{d\theta}, \frac{d\varepsilon_t}{d\theta}, \frac{d^2h_t}{d\theta d\theta'}$ and $\frac{d^2\varepsilon_t}{d\theta d\theta'}$ which are uniformly bounded (we cand demonstrate that the last two matrices are uniformly bounded following by deriving in the recursive expressions (A.2) and (A.3) and following a similar reasoning that employed with the other series).

Therefore, the assumptions of the example 7.1 of Sin and White (1996) are satisfied.

Furthermore, by the ergodic theorem:

$$\frac{B_{s,t}(\theta)}{t-s} = \frac{1}{t-s}\sum_{t=s}^{t}\left[\frac{1}{2h^2{}_t}\frac{dh_t}{d\theta}\frac{dh_t}{d\theta'} + \frac{1}{h_t}\frac{d\varepsilon_t}{d\theta}\frac{d\varepsilon_t}{d\theta'}\right] \longrightarrow E\left[\frac{1}{2h_t^2}\frac{dh_t}{d\theta}\frac{dh_t}{d\theta'} + \frac{1}{h_t}\frac{d\varepsilon_t}{d\theta}\frac{d\varepsilon_t}{d\theta'}\right] \text{ a.s. } (P_0)$$

with uniform convergence on θ. It follows that :

$$B_{s,t}\left(\hat{\theta}_{s,t}\right) = O(t-s) \text{ a.s. } (P_0)$$

because by Proposition 5.1 of Sin and White (1996), $\hat{\theta}_{s.t} \longrightarrow \theta_o$ a.s. (P_0) when t-s $\to \infty$.

The thesis of the theorem is proved following a similar proof to Theorem 4.1 and 4.6 of Muñoz et al. (2001) ⊲

References

Andrews, D. (1987). Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers. *Econometrica,* 55, 1465-1471.

Baillie, R.T. and Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics,* **52**, 91-113.

Bera, A.K. and Higgins, M. (1993). ARCH models: Properties, Estimation and Testing. *Journal of Economic Surveys,* **7**, 305-366.

Bollerslev, T.; Chou, R.Y. and Kroner, K. (1992). ARCH modelling in finance: a review of the theory and empirical evidence. *Journal of Econometrics,* **52**, 5-59.

Davidson, J. (1994). *Stochastic Limit Theory.* Advanced Texts in Econometrics. Oxford University Press.

Lamoureux, G.C. and Lastrapes, W.D. (1990). Persistence in variance, structural change and the GARCH model. *Journal of Business&Economic Statistics,* **8**, 225-234.

Lejeune, B. (1997). Second order pseudo-maximum likelihood estimation and conditional variance misspecification. *University of Liège, ERUDITE and CORE*

Lumsdaine, R. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica,* 64, 575-596.

Miguel, J. and Olave, P. (2002). Adjusting forecast intervals in ARCH-M models. Forthcoming in *Journal of Time Series Analysis*.

Muñoz, L. (2000). *Selección del Modelo y su Base Informativa Óptima en un Contexto Heterocedástico Dinámico: Aplicación a Modelos ARMA-GARCH.* Unpublished Doctoral Dissertation. Faculty of Economic and Business Studies. University of Zaragoza.

Muñoz, L. ; Olave, P. and Salvador, M. (2001). Joint Selection of the Model and the Information Set in Heteroscedastic Dynamic Models. (submitted to *Communications in Statistics: Theory and Methods*).

Phillips, P. (1995). Bayesian Model Selection and Prediction with Empirical Applications (with discussion). *Journal of Econometrics*, **69**, 289-365.

Phillips, P. (1996). Econometric Model Determination. *Econometrica,* 64, 763-812.

Phillips, P. and Ploberger, W. (1994). Posterior Odds Testing for a Unit Root with Data-based Model Selection. *Econometric Theory,* **10**, 774-808.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics,* **6,** 461-464.

Sin, C. and White, H. (1992). Information criteria for selecting possibly misspecified parametric models. *Department of Economics. Discussion Paper 92-47* (University of California, San Diego, CA).

Sin, C. and White, H. (1996). Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics*, **71**, 207-225

Weiss, A. (1986). Asymptotic Theory for ARCH models: estimation and testing. *Econometric Theory ,* 2, 107-131.

Documentos de Trabajo

Facultad de Ciencias Económicas y Empresariales. Universidad de Zaragoza

2002-01: "Evolution of Spanish Urban Structure During the Twentieth Century". Luis Lanaspa, Fernando Pueyo y Fernando Sanz. Department of Economic Analysis, University of Zaragoza.

2002-02: "Una Nueva Perspectiva en la Medición del Capital Humano". Gregorio Giménez y Blanca Simón. Departamento de Estructura, Historia Económica y Economía Pública, Universidad de Zaragoza.

2002-03: "A Practical Evaluation of Employee Productivity Using a Professional Data Base". Raquel Ortega. Department of Business, University of Zaragoza.

2002-04: "La Información Financiera de las Entidades No Lucrativas: Una Perspectiva Internacional". Isabel Brusca y Caridad Martí. Departamento de Contabilidad y Finanzas, Universidad de Zaragoza.

2003-01: "Las Opciones Reales y su Influencia en la Valoración de Empresas". Manuel Espitia y Gema Pastor. Departamento de Economía y Dirección de Empresas, Universidad de Zaragoza.

2003-02: "The Valuation of Earnings Components by the Capital Markets. An International Comparison". Susana Callao, Beatriz Cuéllar, José Ignacio Jarne and José Antonio Laínez. Departament of Accounting and Finance, University of Zaragoza.

2003-03: "Selection of the Informative Base in ARMA-GARCH Models". Laura Muñoz, Pilar Olave and Manuel Salvador. Department of Statistics Methods, University of Zaragoza.